

Neuropsychologie et cognition sociale

P. Allain^{1,2}, J. Fortier¹, J. Besnard¹

¹ LUNAM Université, Université d'Angers, Laboratoire de Psychologie des Pays de la
Loire (EA 4638), 49045, Angers

² Département de Neurologie, CHU Angers, 49033, Angers

Auteur correspondant : Philippe Allain : Ph.allain@chu-angers.fr

RESUME

La notion de cognition sociale fait référence à l'ensemble des aptitudes et expériences émotionnelles et sociales régulant les relations entre les individus et permettant d'expliquer les comportements humains individuels ou en groupe. Elle implique des connaissances sociales, la perception et le traitement de signaux sociaux, ainsi que la représentation des états mentaux. La notion de cognition sociale renvoie donc à de multiples aptitudes. Ce chapitre s'arrête sur plusieurs d'entre-elles, à savoir la théorie de l'esprit, l'empathie, le raisonnement et le jugement moral. Nous proposons une approche conceptuelle de chacune de ces aptitudes en nous appuyant, pour une large part, sur les travaux développés dans le champ de la neuropsychologie humaine. Nous nous arrêtons également sur les soubassements neuro-anatomiques de chacune de ces aptitudes.

MOTS CLES

Cognition sociale, théorie de l'esprit, empathie, raisonnement moral, jugement moral.

INTRODUCTION

La neuropsychologie clinique a plusieurs fois rapporté des dissociations entre troubles cognitifs et troubles du comportement chez les patients atteints de lésions cérébrales, montrant que les changements intervenant dans la personnalité de ces patients ne sont pas toujours réductibles à des déficits cognitifs (inhibition, flexibilité, etc.). C'est en particulier le cas chez les patients présentant des lésions frontales ou incluant les lobes frontaux.

C'est, par exemple, en 1848, le cas Phineas Gage. Ce jeune homme athlétique et vif de 25 ans travaille comme chef d'équipe dans les travaux de construction des voies ferrées dans le nord-est des Etats-Unis. En arrivant près de Cavendish, dans le Vermont, la tâche devient plus ardue. Les couches rocheuses deviennent très dures, nécessitant de faire exploser la roche pour maintenir un tracé rectiligne. Phineas Gage dirige les travaux. Il tasse la poudre au moyen d'une barre de fer. Cet après-midi là, Phinéas Gage est de toute évidence distrait. Il commence à tasser la poudre avant que le sable protecteur soit versé. Une détonation inhabituelle et un sifflement bizarre se font entendre. La roche n'a pas bougé, Phineas Gage est étendu au sol. La barre de fer a atterri à quelques 30 mètres, recouverte de sang et de tissu cérébral. Cette barre de 110 cm de long, 3 centimètres de section et 6 kilos, a pénétré sous la joue gauche de Phineas Gage pour ressortir par le sommet de son crâne. Nonobstant ce traumatisme, Phineas Gage survit. Quelques minutes plus tard, il parle et descend de la charrette qui le mène chez le docteur Harlow, qui le soigne pendant plusieurs mois. Phineas Gage s'en sort sans déficit neurologique. Il parle bien mais a perdu la vue de l'œil gauche. Son caractère évolue. Il devient d'humeur changeante, grossier, inconstant, capricieux. Les remontrances les plus sévères du docteur Harlow sont sans effet. Le statut social de Pineas Gage se modifie. Il ne peut plus être chef de chantier, s'engage dans un parcours chaotique (élevage de chevaux, attraction au cirque Barnum de New-York, conducteur de diligence à Santiago et Valparaiso. Pines Gage décède le 21 mai 1861 (13 ans après l'accident) dans une grande crise d'épilepsie. Antonio et Hanna Damasio

[23] ont ré-analysé des données acquises par le docteur Harlow au milieu du 19^{ème}. Sur la base de photographies du crâne de Phinéas Gage, les auteurs ont reconstruit virtuellement son cerveau pour démontrer qu'il souffrait de lésions de la partie orbitaire du cortex préfrontal, probablement à l'origine de son changement de caractère.

La dissociation rapportée par Eslinger et Damasio [32] à propos du patient EVR constitue une autre illustration remarquable de ce problème. Pour mémoire, EVR, homme marié, père de 2 enfants, sans problème particulier a été, à l'âge de 35 ans, opéré d'un méningiome orbitofrontal. Il n'y a pas eu de complication après l'exérèse chirurgicale. Les lésions sont localisées au niveau du cortex orbitaire et frontal inféro-interne. EVR réalise normalement les tests du bilan neuropsychologique (QI verbal = 129 ; QI performance = 135 ; tests de mémoire, langage, perception visuelle, orientation, construction et fonctions exécutives normaux). Cependant, voilà ce qu'écriront Eslinger et Damasio [32] à propos d'EVR : *"After a 3-month recovery period, he returned to accounting and bookkeeping . . . He soon became involved in a home-building partnership with a former coworker, a man of questionable reputation who had been fired from the company . . . The business failed and he had to declare bankruptcy... Thereafter, he drifted through several jobs. He worked as a warehouse laborer, as a building manager, and as an accountant . . . but was fired from each. Employers complained about tardiness and disorganization, although basic skills, manners, and temper were appropriate. Similar difficulties led to a deterioration of his marital life. . Unable to hold a job and separated from his family, E.V.R. moved in with his parents . . . Employment problems continued . . . He needed 2 hours to get ready for work in the morning, and some days were consumed entirely by shaving and hair-washing. Deciding where to dine might take hours . . . He would drive to each restaurant to see how busy it was... Purchasing small items required in-depth consideration of brands, prices, and the best method of payment. He clung to outdated and useless possessions, refusing to part with dead houseplants, old phone books, six broken*

fans . . .” (pp. 1731-1732). Cette citation montre clairement les difficultés quotidiennes d’EVR, difficultés qui échappent au bilan standard.

Nous avons publié [5] une observation du même type plus récemment. Il s’agit de celle du patient RC (célibataire, 36 ans, 9 ans d’études), victime, en 1978, d’un TC grave. Ce patient a subi une importante lobectomie frontale gauche en 1994 en raison d’abcès cérébraux. Il n’y a pas eu de complication post-opératoire et les tests neuropsychologiques des fonctions instrumentales (langage, praxies et gnosies), intellectuelles (QI WAIS-R verbal = 85 ; QI WAIS-R performance = 104 ; PM 38 = 46/60, percentile 75), mnésiques (Indices de la WMS-R : verbal = 101, visuel = 111, différé = 115 ; courbe des 15 mots de Rey = 6/11/14/14/15 ; courbe des 15 signes de Rey = 5/10/14/15/14 ; Rivearmed Behavioural Memory Test 24/24) sont rapidement devenus normaux. RC réalisait correctement les tests exécutifs (Tour de Londres, Trail Making Test, Stroop, Modified Card Sorting Test, fluences, Test des 6 éléments, estimations cognitives, arrangements de scripts). Seules quelques difficultés apparaissaient dans les tâches engageant l’administrateur central de la mémoire de travail (Résolution problèmes numériques à l’oral, Brown-Peterson, Double-tâche). Son comportement pendant l’examen était normal. Son parcours socio-professionnel post-chirurgie a par contre été compliqué. Il a perdu son emploi, a fait du travail non déclaré, s’est isolé socialement et a fini par retourner vivre chez sa mère.

Ce type d’observations a conduit les neuropsychologues à s’engager, à côté de l’utilisation des tests diagnostiques classiques, dans le développement d’autres types d’outils que l’on peut articuler autour de deux principaux courants de pensée, un courant écologique et un courant de cognition sociale. Le courant écologique a notamment donné naissance à plusieurs batteries ou tests neuropsychologiques, développés dans les années 1990-2000, ayant pour objectif premier une meilleure validité écologique. Ces outils ont surtout été développés dans le monde anglo-saxon et incluent par exemple la Behavioral Assessment of the

Dysexecutive Syndrome [95], une batterie visant à mieux appréhender les difficultés exécutives et comportementale des patients dysexécutifs en vie quotidienne.

Le courant axé cognition sociale est plus récent. Il renvoie aux travaux, actuellement en pleine expansion en neuropsychologie humaine, qui suggèrent l'existence de processus et d'informations spécifiques à la régulation du comportement social et à l'analyse de son propre comportement, sous-tendus particulièrement par les régions orbitaires et/ou médianes des lobes frontaux et dont les perturbations se traduiraient par des troubles du comportement social [91, 48, 60]. De façon plus précise, la cognition sociale renvoie à l'ensemble des aptitudes et expériences émotionnelles et sociales régulant les relations entre les individus et permettant d'expliquer les comportements humains individuels ou en groupe [4]. Elle implique l'acquisition de connaissances sociales, la perception et le traitement des signaux sociaux ainsi que la représentation des états mentaux [42]. Ainsi, la cognition sociale renvoie à plusieurs facettes habituellement scindées en deux dimensions [61] : d'une part, la cognition sociale dite « froide », qui renvoie à la théorie de l'esprit, à l'empathie cognitive et à l'inférence pragmatique, d'autre part la cognition sociale qualifiée de « chaude », qui fait référence à la perception des émotions, à l'empathie affective et aux raisonnements / jugements moraux. Nous nous centrerons plus précisément dans ce chapitre sur les trois concepts les plus étudiés dans la littérature, à savoir la théorie de l'esprit, l'empathie et la capacité à effectuer des raisonnements et jugements moraux.

THEORIE DE L'ESPRIT ET EMPATHIE

La théorie de l'esprit : quelques éléments de définition

Depuis la définition initiale du concept de « théorie de l'esprit » (TDE) par les primatologues Premack & Woodruff [73], c'est dans le champ de la psychologie que se sont développés l'essentiel des travaux visant à en préciser les contours théoriques. Comme le

rappelle Georgieff [41], d'une capacité à prédire les comportements et actions des congénères, la TDE est devenue progressivement l'aptitude à accéder aux états mentaux d'autrui, puis à adopter le point de vue de l'autre. En neuropsychologie, bien que les termes « théorie de l'esprit » renvoient à plusieurs acceptions, ils sont unanimement envisagés comme faisant référence à une aptitude centrale de cognition sociale, renvoyant à des capacités métacognitives dont la mise en œuvre faciliterait l'interaction sociale. Ces capacités méta-représentationnelles seraient permises par un système d'inférences conduisant à l'attribution d'états mentaux, tels que les pensées, intentions, croyances ou sentiments. L'intention initiale de Premack & Woodruff [73] était de déterminer la spécificité humaine de ces aptitudes méta-représentationnelles. Si la question fait toujours l'objet de débats passionnés [e.g., 70 *versus* 21, 7], il semble désormais admis que la différence entre les comportements sociaux démontrés par certains primates et ceux permettant l'interaction sociale propre à l'espèce humaine résiderait dans la distinction entre des processus de « bas niveau », automatiques et immédiats, et des compétences de « haut niveau », contrôlées et accessibles à la conscience, envisagées aussi comme des capacités de « métacognition sociale » [39].

Les mécanismes de bas niveau seraient aspécifiques du genre humain. Ils renvoient au traitement implicite de stimuli sociaux tels que la direction du regard, les actions et postures des congénères. La mise en œuvre de ces processus permettrait la création d'un monde partagé [*a shared world* ; 37] permettant l'interaction avec les autres agents de l'environnement social. Les processus de haut niveau seraient quant à eux associés au développement ontogénétique de facultés cérébrales autorisant une compréhension explicite et volontaire de la situation sociale. Ces habiletés renvoient aux processus contrôlés de la TDE et pourraient permettre de distinguer la cognition sociale animale - qui n'autorise pas l'appréhension de congénères comme des agents intentionnels [92] - de la cognition sociale humaine. Ainsi, les capacités de TDE sont à envisager chez l'être humain comme le résultat de l'interaction entre des processus

automatiques et contrôlés, se distinguant de celles d'autres espèces animales par ces capacités de métareprésentation et de raisonnement permettant d'envisager les autres comme des agents intentionnels. Des habiletés cognitives de haut niveau telles que le langage [8], la mémoire autobiographique [89] ou les fonctions exécutives [1] sous-tendraient en partie ces capacités méta-représentationnelles, même si les liens entre ces fonctions cognitives et les processus contrôlés de TDE restent encore débattus et à étudier. Des perturbations de la TDE ont par exemple été observées indépendamment des perturbations des fonctions exécutives [e.g., 50]. Pour ce qui concerne les processus de bas niveau de la TDE, certaines données récentes de la littérature contestent aussi leur spécificité. En particulier, quelques auteurs font remarquer que des comportements supposés engagés ces processus semblent sous-tendus par des processus cognitifs généraux qui n'interviendraient pas uniquement dans le cadre de la TDE et de la situation d'interaction sociale [concept de « *submentalizing* » ; 51; 81].

La dissociation entre processus automatiques et contrôlés fait écho aux aspects fonctionnels de la TDE définis par Sabbagh et al. [80], qui proposent, chez l'humain, la mise en œuvre simultanée de capacités de décodage et de raisonnement. Cette dissociation renvoie également à celle proposée par d'autres auteurs tels Apperly et Butterfill [6] qui distinguent une TDE *implicite* (processus automatiques) d'une TDE *explicite* (processus contrôlés). D'un point de vue théorique, les processus automatiques et contrôlés semblent également pouvoir être associé aux concepts de Théorie de la théorie et de Théorie de la simulation.

Les conceptions théoriques de la TDE

Les tenants de la Théorie de la théorie (TT) [e.g., 10, 62] proposent que notre capacité à expliquer et prédire notre comportement et celui d'autrui est fondé sur l'utilisation de théories concernant le fonctionnement du monde social et de ses règles. La théorie de la simulation (TS) [e.g., 44, 43] rend compte de la compréhension de la vie mentale par un processus de simulation

imaginative, sans élaboration de réflexion ou de « théorie ». Il existe différentes versions de chacune des théories mentionnées. Dans le cadre de la TT, les partisans de l'approche modulaire conçoivent la genèse des capacités de mentalisation comme un processus de maturation d'un module spécialisé [9], tandis que d'autres auteurs postulent un processus d'élaboration sur la base d'une expérience acquise [62]. Ces propositions ont un point de vue commun : dans les deux cas, la compétence de mentalisation est associée au raisonnement et à la capacité métareprésentationnelle - i.e., de formation des représentations de représentations. Concernant la TS, le postulat commun des différentes versions théoriques renvoie au fait que la capacité à attribuer des états mentaux est plus affaire de simulation que de théorie. L'individu simule le point de vue de l'autre ; le résultat de ce processus conduirait à la compréhension et à la prédiction de son comportement. En cela, la TS a été associée au système des neurones-miroirs.

La découverte des neurones-miroirs a permis d'envisager un lien entre la résonance motrice et les capacités de mentalisation, *via* un processus de simulation [e.g., 40, 78]. Les neurones « miroirs » représentent un type particulier de neurones de l'aire F5 (BA 6), activés lorsque le singe exécute une action, mais également lorsqu'il observe l'expérimentateur effectuer un mouvement similaire. L'action observée semble refléter *en miroir* la représentation motrice de l'action effectuée, d'où la terminologie adoptée par les auteurs. Même si le rôle du système miroir chez l'humain ainsi que l'existence même de neurones miroirs analogues à ceux du singe reste une question encore largement débattue dans la communauté scientifique [65], certains auteurs proposent que la perception et l'imitation « non réalisée » (*covert action*) des mouvements de nos congénères nous permettraient de les comprendre, d'où la terminologie de « théorie motrice de la cognition sociale » [15]. Cette proposition théorique est cependant loin de faire l'unanimité [e.g., 18, 54], compte tenu d'une ambiguïté fréquente dans les publications du groupe de Parme : dans un cas, les neurones miroirs *utilisent* les connaissances motrices, dans l'autre ils les *produisent* [79 *versus* 77]. D'autres auteurs signalent que l'interprétation du

but de l'action permise par le système miroir ne correspond pas à la compréhension des désirs et des croyances.

En résumé, pour faire le lien entre les conceptions théoriques exposées dans cette partie et les définitions de la TDE préalablement mentionnées, la TT suppose une capacité de raisonnement nécessaire pour accéder à la compréhension des comportements et intentions d'autrui et semble donc pouvoir être associée aux processus contrôlés, tandis que la TS semble davantage correspondre à la définition des processus automatiques de décodage social, qui pourraient être permis notamment par le système des neurones-miroirs. Ces deux théories ne sont toutefois pas exemptes de critiques. En particulier, la TT souffre du fait qu'elle semble restreindre le traitement de l'intersubjectivité à un processus réflexif nécessitant un raisonnement inductif, alors que la rapidité des comportements décisionnels en interaction sociale questionne la possibilité d'un tel raisonnement. La TS apparaît également réductrice dans la mesure où les comportements sociaux ne semblent pas uniquement guidés par des processus implicites exempts de tout raisonnement conscient. C'est pourquoi certains auteurs ont proposé des modèles conciliant ces deux théories [64, 57]. Cette relative complexité pour ce qui concerne les postulats théoriques de la TDE va de pair avec une certaine intrication des définitions concernant deux domaines de la cognition sociale : la théorie de l'esprit et l'empathie.

Théorie de l'esprit et empathie

Les termes de « théorie de l'esprit » et d'« empathie » sont parfois employés indistinctement dans la littérature [e.g., 41]. Si le concept de TDE semble recouvrir celui d'empathie dans certaines publications, il semble généralement admis que l'empathie renvoie plus spécifiquement aux processus d'élaboration d'inférences sur les états affectifs des autres, permettant de partager et/ou ressentir leur émotion. Ainsi, Decety [24] suggère clairement de

distinguer la TDE, qui renverrait à la capacité d'attribution d'un état mental à l'autre, de l'empathie, envisagée comme la capacité à éprouver les expériences émotionnelles d'autrui. Auparavant, Decety et Lamm [27] ont proposé de définir l'empathie comme la faculté permettant de comprendre ce que les autres pensent, mais également d'en faire l'expérience, sans confusion avec soi-même. Les capacités d'empathie permettraient donc de partager le ressenti et les émotions des autres. Ces auteurs envisagent également l'empathie sous la dépendance de quatre processus distincts fonctionnant en parallèle : la contagion émotionnelle, la conscience de soi, la prise de perspective subjective de l'autre et les processus de régulation. Il semble possible ici de faire le lien avec les concepts évoqués précédemment pour la TDE : la contagion émotionnelle renvoie à des processus implicites, automatiques, de bas niveau, alors que les processus de régulation sont associés à des processus de haut niveau, dont les fonctions exécutives.

Comme la TDE, l'empathie est donc définie comme un processus engageant plusieurs composantes, ce qui a conduit certains auteurs à distinguer l'empathie *émotionnelle* (également qualifiée de contagion émotionnelle) de l'empathie *cognitive* [67]. La composante émotionnelle (processus de bas niveaux) est directement associée à la « résonance affective », termes utilisés pour désigner le fait que l'observation d'une émotion chez un congénère est susceptible d'induire une émotion identique chez l'observateur, par l'activation de représentations identiques. Ce mécanisme serait étroitement associé au système des neurones-miroirs, point commun supplémentaire entre les concepts de TDE et d'empathie. La composante cognitive de l'empathie impliquerait quant à elle des processus de haut niveau, tels que les fonctions exécutives (inhibition en particulier) mais également la capacité à se distinguer d'autrui, à adopter sa perspective et lui attribuer des pensées et des émotions [27]. Ces processus ne sont pas sans rappeler ceux précédemment évoqués pour la TDE, ce qui peut aussi permettre d'expliquer la confusion parfois retrouvée dans la littérature entre les concepts. La proposition

théorique de Coricelli [22 voir également 16] a pu également renforcer cette confusion terminologique entre TDE et empathie. L'auteur propose en effet d'articuler ces concepts en distinguant une TDE *cognitive*, permettant de faire des inférences sur les états épistémiques des autres sans implication affective ni émotionnelle, d'une TDE *affective*, correspondant à la définition de l'empathie comme la capacité à comprendre le ressenti émotionnel d'autrui. Selon le même modèle, d'autres auteurs ont proposé dans le même temps une dissociation identique pour l'empathie, en distinguant l'empathie *cognitive* de l'empathie *affective* [88], ce qui a également pu contribuer à une certaine confusion terminologique. Selon ces auteurs, l'empathie cognitive renvoie à la capacité d'adopter la perspective d'autrui et comprendre ce qu'il ressent sans toutefois l'éprouver, alors que l'empathie affective renvoie à la capacité à ressentir l'état émotionnel de l'autre. Les termes de « TDE affective » et d'« empathie cognitive » semble donc se recouper et renvoyer à des concepts similaires, ce que les auteurs reconnaissent par ailleurs [30].

En résumé, l'exposé des différents concepts de TDE et d'empathie met en évidence plusieurs points communs : outre le fait que leurs définitions se superposent parfois, ces processus sont multimodaux et semblent mettre en jeu des composantes automatiques et/ou contrôlés pour permettre l'attribution d'états mentaux qui peuvent être cognitifs et/ou affectifs. Néanmoins, d'un point de vue neuro-anatomique, les dissociations proposées semble valide puisque confirmées par des études lésionnelles [voir 86 pour la TDE cognitive *versus* affective] et des travaux en neuroimagerie [voir 87 pour l'empathie émotionnelle *versus* cognitive et 56 pour TDE *versus* empathie]. Ces points seront abordés plus précisément dans la partie suivante.

Les corrélats neuro-anatomiques de la TDE

Les soubassements neuro-anatomiques supposés sous-tendre les processus de TDE ont fait (et font encore) l'objet de nombreux travaux. Le sujet est toujours débattu [voir 85 pour une

revue de la littérature], mais il semble possible néanmoins de dégager des structures cérébrales fréquemment évoquées dans ces études. Par souci de clarté et de lisibilité, nous avons donc scindées ces régions cérébrales en trois groupes distincts, selon la dissociation exposée précédemment qui fait l'objet du plus grand nombre de travaux (TDE cognitive/affective).

Les régions limbiques et para-limbiques

Elles incluent le cortex orbito-frontal, l'amygdale et le gyrus cingulaire antérieur. Ces régions sous-tendraient les compétences associées au versant affectif de la TDE. Ainsi, Eslinger [32] propose que les troubles socio-cognitifs engendrés par des lésions orbito-frontales correspondent à un défaut de ce processus, conduisant à des comportements antisociaux. Il rejoint en cela le point de vue de Damasio [23], qui propose le terme de « sociopathie » acquise pour définir l'ensemble de ces perturbations après lésions orbitaires. Le travail de Stone, Baron-Cohen et Knight [90] corrobore ce point de vue. Les auteurs ont montré que les patients souffrant d'atteinte des régions frontales orbitaires sont moins performants que les sujets porteurs de lésions frontales dorsolatérales à une tâche de « faux pas » nécessitant des réponses affectives. En revanche, ces patients sont tout à fait capables de verbaliser les états émotionnels ou les croyances des protagonistes, contrairement aux malades souffrant de lésions dorsolatérales. Ainsi, cet aspect méta-représentationnel, assimilable au versant « cognitif » de la TDE semble davantage dépendre des structures préfrontales dorsolatérales [22]. A l'appui, Hynes et al. [52] ont démontré l'activation supérieure du cortex orbito-frontal médian [BA 11, 25] lors d'une tâche de TDE affective, en comparaison avec la réalisation d'une épreuve de TDE cognitive.

Les régions préfrontales médianes

De nombreuses études de neuro-imagerie ont démontré l'implication des structures préfrontales médianes dans la réalisation de tâches de TDE. Gallagher, Happé, Brunswick, Fletcher, Frith et Frith [38] rapportent ainsi l'activation du gyrus frontal médian uniquement lors du traitement d'items (des dessins) nécessitant la mobilisation des capacités de TDE cognitive (réponses concernant les croyances ou intentions de protagonistes), en comparaison de dessins neutres [voir aussi 17]. La lecture d'histoires requérant l'attribution d'états mentaux provoque également l'activation spécifique du cortex préfrontal médian, en comparaison d'items ne nécessitant pas la mise en oeuvre de ces processus [93]. Des résultats similaires sont constatés pour d'autres épreuves, comme la vision de formes géométriques évoluant selon des patterns comportementaux laissant supposer une intentionnalité [e.g., 20].

Les régions postérieures

Elles incluent le système pariétal postérieur droit, et plus particulièrement le lobule pariétal inférieur et le sillon temporal supérieur droit. Castelli et al. [20] ont démontré la mise en jeu du sillon temporal supérieur (STS) lors d'une tâche de jugement d'intentionnalité des mouvements des formes géométriques. Les études citées précédemment montrent également une activité au niveau des pôles temporaux, plus marquée au niveau de l'hémisphère gauche [17]. Ces auteurs ont montré que le STS est impliqué dans la compréhension de la causalité et de l'intentionnalité, région également activée lors de la vision de mouvements corporels (e.g., mouvements des lèvres, des yeux, orientation du regard). Saxe et ses collaborateurs [83] ont quant à eux déterminé l'implication de régions de la jonction temporo-pariétale.

En somme, les données de l'imagerie fonctionnelle tendent à confirmer la dissociation proposée entre TDE cognitive et affective. Ainsi, le versant cognitif de la TDE semble soutenu de façon prioritaire par le cortex préfrontal dorso-médian et dorso-latéral, tandis que le versant affectif serait supporté par les régions frontales ventromédianes et orbitaires [e.g., 52,

94]. Plus récemment, Abu-Akel et Shamay-Tsoory [2] proposent d'associer au réseau de TDE cognitive le striatum dorsal, le pôle temporal dorsal et le cortex cingulaire antérieur dorsal, et d'intégrer le striatum ventral, l'amygdale, le pôle temporal ventral et le cortex cingulaire antérieur ventral au réseau de TDE affective. Il faut également souligner que si les soubassements neuro-anatomiques de la TDE cognitive et affective semblent bien identifiés, il n'en est pas de même des processus impliqués dans les aspects implicites de la TDE. L'un des rares travaux à s'être intéressé à cette question a été publié par Kovács et al. [59]. Les auteurs rapportent l'activation de la jonction temporo-pariétal droite et du cortex préfrontal médian lors d'une tâche de TDE implicite.

Les corrélats neuro-anatomiques de l'empathie

Empathie émotionnelle

Comme évoqué précédemment, l'empathie émotionnelle est principalement associée au système de neurones-miroirs. Dans l'un des premiers articles consacrés au sujet, Preston et de Waal [74] ont en effet proposé que l'observation d'un état émotionnel pouvait induire l'activation de réseaux neuronaux responsables d'une expérience émotionnelle analogue chez l'individu, en l'associant au couple perception-action et aux neurones-miroirs. Lors de leur découverte, les neurones-miroirs ont suscité beaucoup d'intérêt et certains auteurs ont pu y voir la base neurobiologique de la cognition sociale. Or, si ce système de neurones apparaît essentiel dans le codage des actions dirigées vers un but, leur rôle dans la perception des émotions reste encore débattu [e.g., 25]. Les premières études en imagerie par résonance magnétique fonctionnelle (IRMf) ont démontré que l'activation cérébrale associée à un état émotionnel se recouvrait avec celle liée à l'observation d'une personne partageant la même expérience [e.g., 19]. Néanmoins, les revues de littérature publiées par la suite semblent indiquer des fluctuations dans l'implication du système des neurones-miroirs dans la perception et l'expression des

émotions, en fonction des informations disponibles dans l'environnement et du contexte [12]. Cela semble également être le cas pour l'empathie cognitive.

Empathie cognitive

Contrairement à l'empathie émotionnelle, l'empathie cognitive impliquerait des processus cognitifs de haut niveau tels que la capacité à se distinguer d'autrui, à adopter sa perspective et à lui attribuer des pensées et des émotions [27]. En premier lieu, être capable d'empathie suppose au préalable d'être en mesure de faire la différence entre soi et les autres. Un prérequis indispensable à la capacité d'empathie réside donc dans la capacité à se distinguer des autres en ayant conscience de soi, rôle qui semble associé à l'intégrité du lobule pariétal inférieur [67]. D'autre part, les travaux dans ce champ de recherche démontrent que le changement de point de vue induit l'activation de régions spécifiques selon la perspective adoptée (s'imaginer dans une situation douloureuse *versus* imaginer un autre individu dans une situation identique), avec l'activation supérieure de la région temporo-pariétale, du cortex cingulaire et du cortex préfrontal médian dans la condition impliquant autrui [53]. Certains auteurs ont également fait remarquer que l'empathie cognitive induit des activations identiques à celles de la TDE, en particulier des processus sous-tendant l'attribution d'une émotion, et que les régions cérébrales recrutées lors d'une épreuve de TDE avec inférence émotionnelle sont identiques à celles mises en œuvre lors des épreuves évaluant l'empathie [86]. Ceci corrobore l'idée qu'empathie cognitive et TDE affective sont deux concepts superposables.

Une grande partie des travaux dans ce champ de recherche a été consacrée à l'étude de la perception de la douleur chez les autres, et démontrent principalement l'activation de la partie antérieure de l'insula, du cortex cingulaire antérieur et de la substance grise périaqueducale. Ces travaux démontrent également des modulations en fonction de divers facteurs sociaux, tels que l'appartenance à un groupe ou le statut de la personne placée dans une situation

douloureuse. Une étude en IRMf fréquemment citée [26] montre ainsi des différences d'activation cérébrale dans les régions associées au traitement de la douleur lorsque l'on fait regarder aux participants des vidéos de personnes porteuses du virus de l'immunodéficience humaine (VIH) qui montrent des expressions de douleur. L'activation est supérieure dans ces régions lorsque l'on indique au sujet que les patients qu'ils voient ont été contaminés par transfusion sanguine (contamination indépendante du sujet), elle est inférieure lorsqu'on leur indique que la contamination résulte de comportements délictueux (prise de drogue avec partage de seringues). Autrement dit, l'intensité de la douleur ressentie, si tant est qu'elle puisse être évaluée par le niveau d'activation cérébrale, apparaît fluctuante et semble dépendre du contexte dans lequel la contamination par le VIH a eu lieu. De la même façon, le contexte social dans lequel survient la situation douloureuse semble avoir un impact sur les régions cérébrales activées. Ainsi, lorsque la situation douloureuse est le résultat d'une action intentionnelle effectuée par un individu (un enfant qui en frappe un autre), les régions qui sous-tendent le raisonnement moral (cortex médian et orbito-frontal, amygdale) sont recrutées, en plus des régions classiquement associées au traitement des informations nociceptives [28]. Ces résultats démontrent ainsi les liens étroits entre les concepts de cognition sociale, et plus spécifiquement ici entre l'empathie et le raisonnement moral. Ces aspects seront abordés plus précisément dans la partie suivante.

RAISONNEMENT ET JUGEMENT MORAL

Comme indiqué plus haut, Godefroy et al. [42] ont récemment rappelé qu'il existait différentes habiletés en cognition sociale, dont le jugement et le raisonnement moral qui ont aussi fait l'objet de nombreux travaux ces dernières décennies. Après quelques rappels conceptuels sur le raisonnement et le jugement moral, nous nous arrêterons sur les principales

données expérimentales et neuroanatomiques disponibles dans la littérature autour de ces aptitudes.

La morale est généralement définie comme l'ensemble des règles qui permettent aux êtres humains de vivre harmonieusement dans le groupe social dont ils font partie. Ces règles sont fondées sur le respect d'autrui et sur l'interdiction de lui faire du mal. Dans l'approche philosophique, une conduite morale peut se définir soit par la valeur morale d'une action (bonne ou mauvaise), soit par ses conséquences ou effets (utilitarisme, pragmatisme), soit par la conformité des valeurs (déontologie); c'est-à-dire selon les intentions ou motivations des individus indépendamment des conséquences de leurs actes. En se basant sur les travaux de Piaget [72], Kohlberg [58] a élaboré un modèle de développement de la morale en 6 étapes partant du raisonnement immature avec un point de vue égocentrique, basé sur l'obéissance et les punitions, pour aller jusqu'à un raisonnement plus mature, basé sur des principes éthiques universels.

Le raisonnement moral

Le raisonnement moral se définit comme la manière dont nous évaluons les relations interpersonnelles et les conventions normatives, ou morales, régulant la vie sociale. Bien que la TDE et le raisonnement moral réfèrent tous deux à une considération des états mentaux d'autrui (relations interpersonnelles), le raisonnement moral se focalise sur la considération de l'intention et de la justification des actions morales, plutôt que sur la compréhension des croyances d'autrui [9].

Afin d'évaluer le raisonnement moral de manière expérimentale, vers la fin des années 1990, Cohen et Greene décident de reprendre les tâches de dilemmes moraux mis en place par Foot en 1967 [34]. Ces tâches sont composées de dilemmes personnels et interpersonnels basés sur le «Trolley Problem». Le dilemme moral personnel, dit de la passerelle («footbridge»), se

compose comme suit : le sujet est confronté à une situation où un tramway est lancé sur une voie et va écraser 5 personnes qui travaillent plus loin sur les rails. Le sujet se trouve sur une passerelle qui domine la voie et d'où il observe la scène. Près de lui, se trouve un homme, un étranger, qui est de forte corpulence. La seule manière de sauver les 5 hommes est qu'il pousse cet homme corpulent pour qu'il chute sur la voie et que son corps arrête le tramway. Bien sûr, cet homme mourra, mais les 5 autres seront sauvés. La question est posée au sujet de savoir s'il déciderait de pousser l'homme se trouvant à côté de lui ou non. Lors d'une tâche avec des sujets sains, 88% des personnes refusent l'idée de pousser l'homme sur la voie [49]. Ils privilégient donc un choix déontologique (respect de la règle de ne pas tuer un autre être humain), plutôt qu'un choix utilitariste (tuer un homme pour en sauver 5).

Lors du dilemme impersonnel («trolley»), cette fois-ci, le sujet est au volant d'un tramway qui approche à grande vitesse d'un aiguillage. Sur les rails qui se dirigent vers la gauche, se trouve un groupe de 5 travailleurs. Sur les rails qui se dirigent vers la droite, se trouve un seul travailleur. S'il ne fait rien, le tramway prendra la voie de gauche et causera la mort des 5 personnes. Le seul moyen d'éviter la mort de ces travailleurs est d'appuyer sur un bouton sur le tableau de bord, ce qui permettra au tramway de prendre la voie de droite, et de causer la mort d'un seul travailleur. Le sujet doit donc décider d'appuyer, ou non, sur le bouton, et donc de ne pas tuer les 5 travailleurs. Lors de ce dilemme, seuls 15% des sujets ne dévient pas le tramway de sa trajectoire [49], tuant donc les 5 travailleurs. Ici les sujets font majoritairement un choix utilitariste (tuer une personne pour en sauver 5), plutôt que déontologique (refuser de choisir de tuer quelqu'un).

Ces dilemmes ont été utilisés avec des patients atteints de pathologies neurodégénératives, et notamment avec des patients atteints de démence frontotemporale avec variante frontale (vfDFT). Les résultats avec le dilemme de la passerelle montrent que la moitié des patients vfDFT choisissent de pousser l'homme sur la voie, contre moins de 20% chez les

sujets sains [49, 63]. Lorsque l'on demande aux patients atteints de vfDFT d'expliquer leurs réponses lors de ce dilemme, ils se justifient de manière rationnelle (sauver le plus de vies possibles) et ne parlent pas d'une quelconque gêne émotionnelle. Ces résultats sont mis en lumière par les performances des patients vfDFT lors de tâches de TDE affectives, évaluées par le Reading the mind in the eyes [11] et d'empathie, où les patients présentent des scores déficitaires [75, 76]. Il existerait donc une forte corrélation entre un déficit d'empathie et de TDE affective chez les patients DFT et leur absence de réaction affective lors d'un dilemme personnel.

Ces différences peuvent s'expliquer par le modèle de Greene et collaborateurs [46, 47] où, selon les auteurs, les réponses lors d'une tâche de raisonnement social de type dilemme seraient soit des réponses émotionnelles automatiques ou des réponses cognitives, plus contrôlées. Plus spécifiquement, les auteurs proposent que les jugements moraux utilitaires (cf : tuer une personne pour en sauver 5) seraient conduits par des traitements cognitifs contrôlés, tandis que les jugements non utilitaires, ou jugement déontologique (cf. : ne pas choisir de tuer une personne), seraient conduits par des réponses émotionnelles automatiques. Cette dissociation pourrait donc être à la base du manque émotionnel des patients vfDFT. Leurs performances lors d'une tâche d'empathie et de TDE affective étant significativement inférieures à celles de patients contrôles, selon le modèle de Greene, les jugements déontologiques seraient déficitaires. Cela expliquerait donc en quoi les patients atteints de vfDFT ont une tendance significativement plus grande que les contrôles à ne pas hésiter à utiliser l'homme obèse dans le dilemme personnel.

Les jugements moraux

Les jugements moraux sont définis comme les jugements effectués sur les comportements des autres personnes comme étant bien ou mal, mais aussi sur leur façon de se comporter avec les autres personnes [99].

Afin d'évaluer les jugements moraux, Young, Nichols et Saxe [98] ont développé un scénario de «Moral luck», présentant deux situations différentes : « *Grace et sa collègue de travail visitent une usine de produits chimiques. Grace s'arrête pour prendre un café pour elle et pour sa collègue. A côté se trouve un container de sucre. Le contenu est marqué comme toxique, donc Grace pense que le contenu est toxique. Elle met une cuillère du contenu dans le café de sa collègue et rien dans le sien. Quand sa collègue boit le café, rien de mauvais n'arrive.* » Ce premier scénario montre une intention de blesser, mais sans conséquence. Le scénario alternatif est le suivant : « *Un container de poison se trouve à côté du café. Il est marqué sur le container « sucre », donc Grace pense qu'il s'agit de sucre. Elle met une cuillère dans le café de sa collègue. Sa collègue boit le café et meurt.* ». Dans ce scénario un accident arrive, une conséquence tragique due à une fausse croyance, mais sans intention négative. La plupart des études montre que les sujets accordent plus d'importance morale aux croyances et intentions des personnes qu'aux résultats de leurs actions [97]. Ils jugent comme étant moralement pire d'essayer de blesser quelqu'un sans y arriver que de blesser quelqu'un accidentellement [98]. Lors du jugement des actions de Grace, les zones cérébrales activées sont exactement les mêmes que celles activées lors de raisonnements sur les états mentaux [55, 71, 82, 96]: principalement la jonction temporo-pariétale droite pour le jugement d'une action ayant des conséquences graves.

Ces activations des zones correspondant aux TDE montrent que les sujets ne jugeraient pas uniquement une action sur ses conséquences, mais bien sur un ensemble de variables avec, au premier plan, les états mentaux de la personne effectuant l'action (croyances, intentions). Ils

modifieraient donc leur évaluation des états mentaux de Grace (ce que Grace croît au moment de l'action) en fonction des conséquences de l'action.

Un autre axe de recherche sur les jugements moraux concerne les jugements effectués sur des transgressions de normes morales ou conventionnelles. Blair [14] définit les transgressions morales comme des transgressions ayant des conséquences négatives pour les droits et le bien-être des personnes, alors que les transgressions conventionnelles sont définies comme des violations des comportements classiques, d'uniformités, structurants les interactions sociales dans un système. Afin de tester ces différents jugements, Blair [13] a construit une tâche [basée sur les travaux de Dewey, 29] se composant de 18 petites histoires (9 histoires morales, 9 histoires conventionnelles), prenant toutes places dans un contexte scolaire pour enfant. L'exemple le plus utilisé pour la transgression morale est celui d'un enfant entrant dans une salle de classe et détruisant un piano. Pour la transgression conventionnelle, l'exemple classique est celui d'un enfant allant à l'école vêtu d'une jupe.

Pour évaluer comment les sujets jugeaient cette transgression, il leur était demandé, dans un premier temps, de juger si cette action était bien ou mal (est-elle dans la norme morale ou conventionnelle ?) Si la réponse est «mal», des questions supplémentaires leurs étaient alors posées, comme par exemple : Sur une échelle de 1 à 10, à quel point est-ce bien/mal (quantifie la gravité subjective de la transgression) ? Pourquoi est-ce bien/mal (identifier les théories sur l'acte) ? Les résultats montrent un très grand consensus, les sujets jugeant toujours une transgression conventionnelle comme plus permissive qu'une transgression morale. De plus, les transgressions morales sont jugées significativement plus sérieuses que les transgressions conventionnelles. En 2006, Lough et al. [60] ont réutilisé cette tâche avec des patients vDFT . Les résultats ont montré que les patients ne faisaient plus de différence entre les transgressions morales et conventionnelles. Elles étaient jugées toutes deux aussi graves. De plus, ils identifiaient significativement moins de violation de normes que les sujets contrôles.

Récemment, Ehrlé, Henry, Pesa et Backchine [31] ont repris la tâche de Blair sur les jugements moraux et conventionnels pour l'appliquer à une population française. Dans cette adaptation, les histoires ne se déroulent plus dans des écoles mais dans des lieux plus en lien avec le monde adulte (entreprise, etc.). De plus, un groupe de scènes contrôles a été ajouté afin que les sujets ne jugent pas toutes les histoires comme des transgressions. Avec cette adaptation, Ehrlé et al. [31] ont pu démontrer des déficits dans les jugements moraux et conventionnels chez un patient atteint de vDFT, mais sans perturbation de la cognition froide (TDE cognitive) et sans trouble exécutif. Ce test permet donc de tester la cognition sociale chez les patients neurologiques, en particulier lorsqu'ils présentent des troubles du comportement avec des performances cognitives peu dégradées.

Neuroanatomie du cerveau « moral »

Les aspects de hauts niveaux de la cognition sociale, tels que le raisonnement moral, le jugement moral et la prise de décision semblent se reposer grandement sur les régions préfrontales. Une aire cérébrale semble particulièrement importante : il s'agit du gyrus frontal médial. Ce gyrus semble être indispensable à l'intégration des émotions dans la prise de décision [69]. Il semble également jouer un rôle dans les TDE et dans d'autres fonctions cognitives relevant du jugement moral [voir 35]. En ce qui concerne les autres aires cérébrales impliquées, le cortex cingulaire antérieur semble très communément activé lors d'expérience émotionnelles (dilemme par exemple). Le cortex frontal dorsolatéral, ainsi que les lobes pariétaux, semblent également être impliqués dans les jugements moraux impersonnels [45, 47].

Dans leur synthèse, Moll et collaborateurs [66] notent un fort consensus autour de plusieurs régions impliquées dans des tâches de jugement moral, dont les cortex frontaux orbitaux et frontopolaires, le cortex préfrontal ventromédian, ainsi que la jonction temporo-pariétale et les gyri angulaires et cingulaire postérieur.

En résumé, il semble que le raisonnement et le jugement moral sont deux composantes de cognition sociale qui partagent un substrat anatomique similaire. De plus, il existe de forts liens entre les TDE et les jugements et raisonnements moraux. Nous avons en effet vu les liens existant entre TDE affective et raisonnement moral personnel et entre le substratum anatomique du jugement moral et des TDE. Il semble donc nécessaires d'évaluer les différentes facettes de la cognition sociale évoquées ici, et ne pas se contenter d'un seul test d'évaluation de la cognition sociale. Il n'existe pas une, mais des cognitions sociales.

CONCLUSION GENERALE

Ce chapitre se donnait pour double objectif de présenter les définitions des concepts de cognition sociale parmi les plus étudiés dans la littérature, tout en présentant l'actualité des connaissances sur les corrélats neuro-anatomiques associés. Cette revue de la littérature permet de mettre en lumière de fortes relations entre les processus, que ce soit pour les définitions, qui se superposent parfois, ou les structures cérébrales qui les sous-tendent. Malgré l'avancée spectaculaire des connaissances sur le sujet, plusieurs points méritent d'être soulevés.

Premièrement, il reste encore à déterminer la spécificité des processus cognitifs associés aux compétences de cognition sociale. Comme nous l'avons évoqué pour la théorie de l'esprit, les études récentes, qui ont conduit à l'émergence du concept de « *submentalizing* » [51, 81] tendent à démontrer que certains aspects de la cognition sociale peuvent être opérationnalisés de manière très rapide, automatique, sans contrôle conscient de la part du sujet. Ces processus ne seraient pas spécifiquement mis en œuvre lors de la situation d'interaction sociale et donc aspécifiques des compétences de cognition sociale. Pour les processus contrôlés, il semble également difficile d'isoler des compétences spécifiques de cognition sociale de fonctions de haut niveau comme le langage, les fonctions exécutives ou la mémoire. Ce constat a récemment

conduit certains auteurs à proposer de redéfinir certains concepts de cognition sociale, comme celui de la théorie de l'esprit [84].

Deuxièmement, il apparaît que la définition originelle de la cognition sociale implique la capacité à ajuster son comportement en situation d'interaction sociale. Sur la base de cette définition, les auteurs sont nombreux à associer les compétences de cognition sociale à la faculté de se comporter de façon appropriée en situation d'interaction [e.g., 3, 36]. Le corollaire légitime de ce postulat prévoit qu'un déficit de ces capacités méta-représentationnelles est susceptible d'induire des troubles du comportement social. Dans les faits, peu de travaux s'intéressent finalement à valider cette hypothèse. Il reste donc encore à démontrer en quoi les perturbations des processus de cognition sociale rendent effectivement compte de déficits de régulation en situation d'interaction, en recherchant systématiquement, notamment au moyen de questionnaires, des relations entre les perturbations sociocognitives et des troubles du comportement en situation d'interaction sociale. L'autre alternative est de penser de nouvelles méthodologies d'évaluation de la cognition sociale, qui pourraient permettre d'évaluer la capacité des sujets à analyser des situations d'interaction réelles. Une grande partie des travaux portant sur l'étude des processus de cognition sociale implique en effet de l'imagerie fonctionnelle, situation qui limite par définition les possibilités d'évaluer réellement les processus à l'œuvre en situation d'interaction sociale. Par ailleurs, quelques travaux menés dans la pathologie démontrent qu'il peut exister des dissociations entre des compétences préservées de TDE et des troubles du comportement sociaux [e.g., 13, 68]. Ces aspects renforcent une fois encore l'importance cruciale d'une meilleure prise en compte des conditions d'évaluation de ces processus, qui permettent bien souvent de répondre à des questions théoriques sans évaluer empiriquement les processus à l'œuvre lors de situations d'interactions humaines.

Bibliographie

1. Aboulafia-Brakha T, Christe B, Martory MD, Annoni JM. Theory of mind and executive functions: A systematic review of group studies in neurology. *J Neuropsychol* 2011; 5: 39-55.
2. Abu-Akel A, Shamay-Tsoory S. Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia* 2011; 49: 2971-84.
3. Adolphs R. Cognitive neuroscience of human social behaviour. *Nat Rev Neurosci* 2003; 4: 165-78.
4. Allain P, Aubin G, Le Gall D, eds. *Cognition sociale et neuropsychologie*. Marseille : Solal, 2012.
5. Allain P, Etcharry-Bouyx F, Le Gall D. A case study of selective impairment of the central executive component of working memory after a focal frontal lobe damage. *Brain Cog* 2001; 45: 21-43.
6. Apperly IA, Butterfill SA. Do humans have two systems to track beliefs and belief-like states? *Psychol Rev* 2009; 116: 953-70.
7. Arruda CT, Povinelli DJ. Chimps as secret agents. *Synthese* 2015; 1-30.
8. Astington JW, Jenkins JM. A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychol* 1999; 35: 1311-20.
9. Baird JA, Astington JW. The role of mental state understanding in the development of moral cognition and moral action. *New Directions Child Adol Develop* 2004; 103: 37-49.
10. Baron-Cohen S. *Mindblindness. An essay on autism and theory of mind*. Cambridge: MIT Press, 1995.
11. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatr* 2001; 42: 241-51.
12. Bernhardt BC, Singer T. The neural basis of empathy. *An Rev Neurosci*, 2012; 35: 1-23.
13. Besnard J. *Contributions à l'étude des phénomènes de dépendance à l'environnement chez les patients cérébrolésés frontaux*. Thèse de doctorat non publiée. Université d'Angers, 2009.
14. Blair RJR. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition* 1995; 57: 1-29.
15. Blakemore SJ, Decety J. From the perception of action to the understanding of intention. *Nat Rev Neurosci* 2001; 2: 561-67.
16. Brothers L, Ring B. A neuroethological framework for the representation of minds. *J Cog Neurosci* 1992; 4: 107-18.
17. Brunet E, Sarfati Y, Hardy-Baylé MC, Decety J. A PET investigation of the attribution of intentions with a non-verbal task. *NeuroImage* 2000; 11: 157-66.
18. Carpendale JIM, Lewis C. Mirroring cannot account for understanding action. *Behav Brain Sci* 2008; 31: 23-4.

19. Carr L, Iacoboni M, Dubeau MC, et al. Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proc Nat Acad Sci* 2003; 100: 5497-502.
20. Castelli F, Happé F, Frith U, Frith C. Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* 2000; 12: 314-25.
21. Clayton NS, Dally JM, Emery NJ. Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philos T R Soc B* 2007; 362: 507-22.
22. Coricelli G. Two-levels of mental states attribution: From automaticity to voluntariness. *Neuropsychologia* 2005; 43: 294-300.
23. Damasio AR. *L'erreur de Descartes: La raison des émotions*. Paris : Odile Jacob, 1995.
24. Decety J. A social cognitive neuroscience model of human empathy. In : Harmon-Jones E, Winkielman P, eds. *Social neuroscience: Integrating biological and psychological explanations of social behavior*. New York: Guilford Press, 2007: 246-70.
25. Decety J. To what extent is the experience of empathy mediated by shared neural circuits? *Emotion Rev* 2010; 2: 204-7
26. Decety J, Echols SC, Correll J. The blame game: The effect of responsibility and social stigma on empathy for pain. *J Cogn Neurosci* 2009; 22: 985-97.
27. Decety J, Lamm C. Human empathy through the lens of social neuroscience. *Scientific World J* 2006; 6: 1146-63.
28. Decety J, Michalska KJ, Akitsuki Y. Who caused the pain? A functional MRI investigation of empathy and intentionality in children. *Neuropsychologia* 2008; 46: 2607-14.
29. Dewey M. Living with Asperger's syndrome. In : Frith U, ed. *Autism and Asperger syndrome*. Cambridge: Cambridge University Press, 1991: 184-206.
30. Dvash J, Shamay-Tsoory SG. Theory of mind and empathy as multidimensional constructs: Neurological foundations. *Topics Lang Disord* 2014; 34: 282-95.
31. Ehrlé N, Henry A, Pesa A, Backchine S. Présentation d'une batterie d'évaluation des fonctions sociocognitives chez des patients atteints d'affections neurologiques: Application dans la démence frontale. *Geriatr Psychol Neuropsychiatr Vieil* 2011; 9: 117-28.
32. Eslinger PJ. Neurological and neuropsychological bases of empathy. *European Neurol*, 1998; 39: 193-99.
33. Eslinger PJ, Damasio AR. Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology* 1995; 35: 1731-41.
34. Foot P, ed. *Theories of ethics*. Oxford : Oxford University Press, 1977, 130-31.
35. Frith U. Mind blindness and the brain in autism. *Neuron* 2001; 32: 969-79.
36. Frith CD. Social cognition. *Philos T R Soc B* 2008; 363: 2033-39.
37. Frith CD, Frith U. Social cognition in humans. *Curr Biol* 2007; 17: 724-32.
38. Gallagher HL, Happé F, Brunswick N, Fletcher PC, Frith U, Frith CD. Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 2000; 38: 11-21.

39. Gallese V. Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition. *Philos T R Soc B* 2007; 362: 659-69.
40. Gallese V, Goldman A. Mirror neurons and the simulation theory of mind-reading. *Trends Cog Sci* 1998; 2: 493-501.
41. Georgieff N. Penser l'autre, être pensé par l'autre: Intérêts de la notion de théorie de l'esprit pour la psychopathologie. *Confront Psychiat* 2005; 46: 127-42.
42. Godefroy O, Jeannerod M, Allain P, Le Gall D. Lobe frontal, fonctions exécutives et contrôle cognitif. *Rev Neurol* 2008; 164: 119-27.
43. Goldman AI. In defense of the simulation theory. *Mind Lang* 1992; 7: 104-19.
44. Gordon RM. "Radical" simulationism. In : Carruthers P, Smith PK, eds. *Theories of theories of mind*. Cambridge: Cambridge University Press, 1996: 11-21.
45. Greene J, Haidt J. How (and where) does moral judgment work? *T Cog Sci* 2002; 6: 517-23.
46. Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 2004; 44: 389-400.
47. Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. *Science* 2001; 293: 2105-108.
48. Gregory C, Lough S, Stone V, Erzinclioglu S, Martin L, Baron-Cohen S. Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain* 2002; 125: 752-64.
49. Hauser M, Cushman F, Young L, Kang-Xing Jin R, Mikhail J. A dissociation between moral judgments and justifications. *Mind Lang* 2007; 22: 1-21.
50. Havet-Thomassin V, Allain P, Aubin G, Etcharry-Bouyx F, Bernat C, Le Gall D. Adhérence cognitive et déficit d'intentionnalité: Présentation de deux cas dissociés. *Rev Neuropsychol* 2006; 16: 183-215.
51. Heyes CM. Submentalizing: I'm not really reading your mind. *Perspectives Psychol Sci* 2014; 9: 131-43.
52. Hynes CA, Baird AA, Grafton ST. Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia* 2006; 44: 374-83.
53. Jackson PL, Brunet E, Meltzoff AN, et al. Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia* 2006 ; 44 : 752-61.
54. Jacob P, Jeannerod M. The motor theory of social cognition: A critique. *Trends Cog Sci* 2005; 9: 21-5.
55. Jenkins AC, Mitchell JP. Mentalizing under uncertainty: Dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cereb Cortex* 2010; 20: 404-10.
56. Kanske P, Böckler A, Trautwein FM, Singer T. Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and theory of mind. *NeuroImage* 2015; 122: 6-19.

57. Keysers C, Gazzola V. Integrating simulation and theory of mind: From self to social cognition. *Space* 2007; 8: 108-14.
58. Kohlberg L. *The psychology of moral development: The nature and validity of moral stages* (Vol. 2). San Francisco : Harper & Row, 1984.
59. Kovács ÁM, Kühn S, Gergely G, Csibra G, Brass M. Are all beliefs equal? Implicit belief attributions recruiting core brain regions of theory of mind. *PLoS One* 2014; 9: e106558.
60. Lough S, Kipps CM, Treise C, Watson P, Blair JR, Hodges JR. Social reasoning, emotion and empathy in frontotemporal dementia. *Neuropsychologia* 2006; 44: 950-58.
61. McDonald S. Impairments in social cognition following severe traumatic brain injury. *J Int Neuropsychol Soc* 2013; 19: 231-46.
62. Meltzoff AN. Origins of theory of mind, cognition and communication. *J Com Dis* 1999; 32: 251-69.
63. Mendez MF, Anderson E, Shapira JS. An investigation of moral judgement in frontotemporal dementia. *Cog Behavioral Neurol* 2001; 18: 193-97.
64. Mitchell JP. The false dichotomy between simulation and theory-theory: The argument's error. *Trends Cogn Sci* 2005; 9: 174-79.
65. Molenberghs P, Cunnington R, Mattingley JB. Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neurosci Biobehav Rev* 2009; 33: 975-80.
66. Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J. The neural basis of human moral cognition. *Nat Rev Neurosci* 2005; 6: 799-809.
67. Narme P, Mouras H, Loas G et al. Vers une approche neuropsychologique de l'empathie. *Rev Neuropsychol* 2010; 2: 292-98.
68. Ozonoff S, Pennington BF, Rogers SJ. Executive function deficits in high-functioning autistic individuals: Relationships to theory of mind. *J Child Psychol Psychiat* 1991; 32: 1081-105.
69. 91. Partiot A, Grafman J, Sadato N, Wachs J, Hallett M. Brain activation during the generation of nonemotional and emotional plans. *Neuroreport* 1995; 6: 1397-1400.
70. Penn DC, Povinelli DJ. On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind". *Proc Natl Acad Sci B* 2007; 362: 731-44.
71. Perner J, Aichhorn M, Kronbichler M, Staffen W, Ladurner G. Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neurosci* 2006; 1: 245-58.
72. Piaget J. *The moral judgement of the child*. New York: Simon and Schuster, 1997.
73. Premack D, Woodruff G. Does the chimpanzee have a «Theory of mind»? *Behav Brain Sci* 1978; 4: 515-26.
74. Preston SD, de Waal FB. Empathy: It's ultimate and proximate bases. *Behav Brain Sci* 2002; 25: 1-20.
75. Rankin KP, Gorno-Tempini ML, Allison SC et al. Structural anatomy of empathy in neurodegenerative disease. *Brain* 2006; 129: 2945-56.

76. Rankin KP, Kramer JH, Miller B. Patterns of cognitive and emotional empathy in frontotemporal lobar degeneration. *Cog Behav Neurol* 2005; 18: 28-36.
77. Rizzolatti G, Craighero L. The mirror-neuron system. *An Rev Neurosci* 2004; 27: 169-92.
78. Rizzolatti G, Fabbri-Destro M. The mirror system and its role in social cognition. *Current Opinion Neurobiol* 2008; 18: 179-84.
79. Rizzolatti G, Fogassi L, Gallese V. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2001; 2: 661-70.
80. Sabbagh MA, Moulson MC, Harkness KL. Neural correlates of mental state decoding in human adults: An event-related potential study. *J Cog Neurosci* 2004; 16: 415-26.
81. Santiesteban I, Shah P, White S, Bird G, Heyes C. Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bul Rev* 2014; 22: 844-49.
82. Saxe R, Kanwisher N. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* 2003; 19: 1835-42.
83. Saxe R, Xiao DK, Kovacks G, Perret DI, Kanwisher N. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 2004; 42: 1435-46.
84. Schaafsma SM, Pfaff DW, Spunt RP, Adolphs R. Deconstructing and reconstructing theory of mind. *Trends Cog Sci* 2015; 19: 65-72.
85. Schurz M, Radua J, Aichhorn M, Richlan F, Perner J. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev* 2014; 42: 9-34.
86. Shamay-Tsoory SG, Aharon-Peretz J. Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia* 2007; 45: 3054-67.
87. Shamay-Tsoory SG, Aharon-Peretz J, Perry D. Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 2009; 132: 617-27.
88. Shamay-Tsoory SG, Tomer R, Goldsher D, Berger BD, Aharon-Peretz J. Impairment in cognitive and affective empathy in patients with brain lesions: Anatomical and cognitive correlates. *J Clinical Exp Neuropsychol* 2004; 26: 1113-27.
89. Spreng RN, Mar RA, Kim AS. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *J Cog Neurosci* 2009; 21: 489-510.
90. Stone VE, Baron-Cohen S, Knight RT. Frontal lobe contributions to theory of mind. *J Cog Neurosci* 1998;10: 640-56.
91. Stuss DT, Gallup GG, Alexander MP. The frontal lobes are necessary for theory of mind. *Brain* 2001; 124: 279-86.
92. Tomasello M. *The cultural origins of human cognition*. Cambridge : Harvard University Press, 1999.
93. Vogeley K, Bussfeld P, Newen A, et al. Mind reading: Neural mechanisms of theory of mind and selfperspective. *NeuroImage* 2001; 14: 170-81.

94. Völlm BA, Taylor AN, Richardson P, et al. Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage* 2006; 29: 90-8.
95. Wilson BA, Alderman N, Burgess PW, Emslie H, Evans, J. *Behavioural Assessment of the Dysexecutive Syndrome*. Bury St Edmunds : Thame Valley Test Company, 1996.
96. Young L, Camprodon JA, Hauser M, Pascual-Leone A, Saxe R. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc Natl Acad Sci* 2010; 107: 6753-58.
97. Young L, Cushman F, Hauser M, Saxe R. The neural basis of the interaction between theory of mind and moral judgment. *Proc Natl Acad Sci* 2007; 104: 8235-40.
98. Young L, Nichols S, Saxe R. Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Rev Philos Psychol* 2010; 1: 333-49.
99. Young L, Waytz A. Mind attribution is for morality. In : Baron-Cohen S, Lombardo M, Tager-Flusberg H, eds. *Understanding other minds: Perspectives from developmental social neuroscience*. Oxford: Oxford University Press, 2013, 93-103.